

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-13982

(P2001-13982A)

(43) 公開日 平成13年1月19日 (2001.1.19)

(51) Int.Cl.<sup>7</sup>

G 1 0 L 13/06

13/08

識別記号

F I

G 1 0 L 5/04

3/00

テーマワード (参考)

F 5 D 0 4 5

H 9 A 0 0 1

審査請求 未請求 請求項の数 6 O L (全 23 頁)

(21) 出願番号 特願平11-219216

(22) 出願日 平成11年8月2日 (1999.8.2)

(31) 優先権主張番号 特願平11-122718

(32) 優先日 平成11年4月28日 (1999.4.28)

(33) 優先権主張国 日本 (J P)

(71) 出願人 000004329

日本ビクター株式会社

神奈川県横浜市神奈川区守屋町3丁目12番地

(72) 発明者 和田 祐司

神奈川県横浜市神奈川区守屋町3丁目12番地

日本ビクター株式会社内

(74) 代理人 100083806

弁理士 三好 秀和 (外9名)

Fターム (参考) 5D045 AA08 AA09 AA11

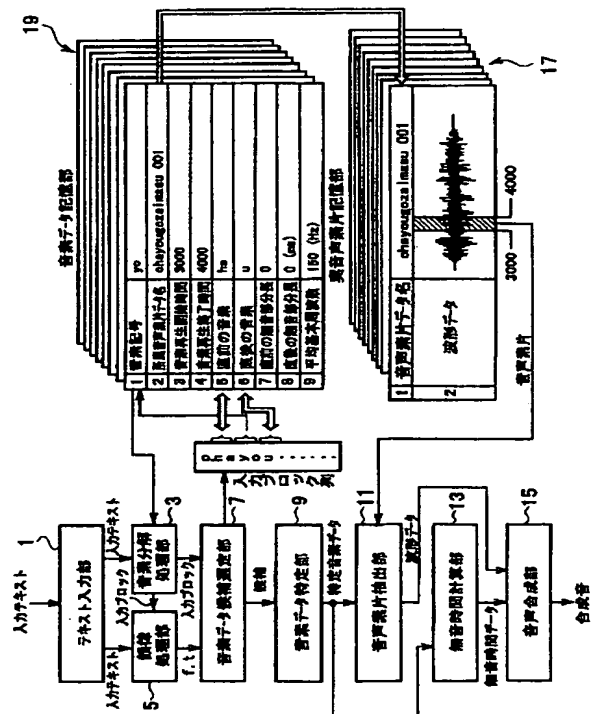
9A001 HH18 HZ30

(54) 【発明の名称】 音声合成装置

(57) 【要約】

【課題】 音声素片の接続部分のノイズを低減するとともに、録音された実音声の特徴を備えた高品質の合成音出力が可能な音声合成装置を提供する。

【解決手段】 音素データ記憶部19の各保存領域には、「音素記号」、「所属音声素片データ名」、「直前の音素」、「直後の音素」が予め登録され、実音声素片記憶部17には、実音声を記憶した波形データが音声素片データ名とともに蓄積して記憶される。音素データ特定部9は、入力ブロック列のなかの互いに隣接する「入力ブロック」の組と、音素データ記憶部19の「音素記号」および「直前の音素」（あるいは、「直後の音素」）とを照合することにより、各入力ブロックに対して隣接音声環境が最適な音素データを特定し、音声素片抽出部11は特定音素データの「所属音声素片データ名」を基に実音声素片記憶部17から音声素片を抽出し、音声合成部15は抽出された音声素片を順次接続し合成音を出力する。



## 【特許請求の範囲】

【請求項 1】 音声波形が予め記憶された波形記憶手段と、  
該波形記憶手段の音声波形に含まれる第 1 の音声素片に対して付与された第 1 のラベルと、当該第 1 の音声素片に隣接する音声素片に対して付与された第 2 のラベルと、当該第 1 の音声素片の記憶場所情報とを含む音素データが予め記憶される音素データ記憶手段と、  
音声合成指令としての入力音素記号列と該入力音素記号列に隣接する隣接入力音素記号列とからなる組と、前記音素データに含まれる第 1 のラベルおよび第 2 のラベルの組とを照合して、当該入力音素記号列に対する音素データを特定する音素データ特定手段と、  
前記波形記憶手段のなかの前記特定された音素データの示す記憶場所に記憶された音声素片を抽出する音声素片抽出手段と、  
該音声素片抽出手段で抽出された音声素片を所定の順序で接続し出力する音声合成手段とを有する音声合成装置。

【請求項 2】 請求項 1 記載の音声合成装置において、前記音素データには、前記第 1 の音声素片と該第 1 の音声素片に隣接する音声素片との間の状態に係る時間情報が付与され、該時間情報に基づいて音声素片の接続部分の状態を設定する時間設定手段を有することを特徴とする音声合成装置。

【請求項 3】 請求項 2 記載の音声合成装置において、前記時間設定手段は、前記音素データに、前記第 1 の音声素片と該第 1 の音声素片の前後に隣接する各音声素片との間の無音部分の時間情報がそれぞれ付与されているとき、前記第 1 の音声素片の後の無音部分の時間情報と、当該第 1 の音声素片の後に接続される音声素片の前の無音部分の時間情報の少なくとも一方に対し重み付けを行って、当該各音声素片を接続する部分に設ける無音時間を算出することを特徴とする音声合成装置。

【請求項 4】 請求項 1 ないし請求項 3 のいずれかに記載の音声合成装置において、前記入力音素記号列に対して特定された音素データが、当該入力音素記号列の前又は後ろに隣接する入力音素記号列に対して特定されることを規制する規制手段を有することを特徴とする音声合成装置。

【請求項 5】 請求項 1 ないし請求項 4 のいずれかに記載の音声合成装置において、前記音声合成手段により出力される音声素片の出力タイミングに係る情報を生成するタイミング情報生成手段を具備することを特徴とする音声合成装置。

【請求項 6】 請求項 5 記載の音声合成装置において、前記タイミング情報生成手段は、前記音声合成手段により出力される音声素片に当該音声素片よりも短い音声が含まれるときには、当該音声との同期をとるための情報を前記出力タイミングに係る情報に含ませることを特徴

とする音声合成装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、音声素片の接続部分のノイズを低減するとともに、録音された音声の特徴を備えた高品質の合成音出力が可能な音声合成装置および合成音に同期させて、顔画像における口の形状を変化させることのできる音声合成装置に関する。

## 【0002】

【従来の技術】近年、人の実音声を録音して保存したもののから合成音を生成する録音合成の技術が、音声の特徴の再現性の点から注目されている。

【0003】従来の録音合成を利用した音声合成装置にあっては、予め実音声を録音し、例えば、「あ」や「い」等の 50 音を構成する音声等の所定の音声単位ごとに区切って多数の音声素片としたものが用意され、この各音声素片は、音声素片の種類を表す音素記号、音声の高低を決める基本周波数および再生時間長が付与されてデータベースとして蓄積保存される。

【0004】テキストから音声を合成する際には、入力された日本語テキストは、先ず単語辞書を基にした音素分解処理により音素記号に分解され、その得られた各音素記号に対して、所定のアクセントパターン表および再生時間長規則を基にした韻律処理により基本周波数と再生時間長とが割り当てられる。そして、音素記号、基本周波数および再生時間長と、上述のデータベースのなかの音素記号、基本周波数および再生時間長とが照合され、音素分解処理により得られた音素記号に対して音声素片が選択される。そして、選択された音声素片が入力されたテキストの順番で接続されて合成音声が生じられる。

【0005】また、音声合成の技術が注目されるなかにあつて、その一方では、音声解析や CG (Computer Graphics) の技術の加速的な進歩により、入力された音声波形を解析しながら、実際に人が音声をしゃべっているように口の形状が変化する顔画像を生成する画像処理技術が実現段階に入っている。このような技術は、例えば、事件の現場から電話により伝えられた言葉を、CG で制作されたアナウンサにしゃべらせるというような用途などに利用される。

【0006】このような用途において、顔画像における口の形状をリアルタイムに変化させるためのパラメータを生成するためには、コンピュータシステムなどを利用して、入力された音声波形の解析処理が行われる。

## 【0007】

【発明が解決しようとする課題】しかしながら、上記従来の音声合成装置では、合成音声の生成のための音声素片を記憶するにあたっては、実音声において隣接する音声素片に関係なく互いに分離され、一方、音声素片を合成するにあたっては、入力された音素記号等と、データ

ベースの各音声素片に対応づけられた音素記号等とを照合することにより音声素片を選択し、その選択された音声素片を単純に接続する方法が採られていたため、生成される合成音は、音声素片の接続部分に生ずるノイズが高く、録音された実音声の特徴が損われてしまう場合があった。

【0008】また、上記のように、顔画像における口の形状を音声に同期して変化させるためには、処理能力の高いコンピュータシステムによる複雑で膨大な音声解析処理が必要であることから、このような音声解析処理を行わずに、顔画像における口の形状を音声と同期したタイミングで変化させたいとの要望が多くなっている。

【0009】そこで本発明は、上記従来の問題点に鑑みてなされたものであり、その目的とするところは、音声素片の接続部分のノイズを低減するとともに、録音された実音声の特徴を備えた高品質の合成音出力が可能な音声合成装置を提供することにある。

【0010】また、本発明の他の目的とするところは、合成音に同期させて、顔画像における口の形状を変化させることのできる音声合成装置を提供することにある。

#### 【0011】

【課題を解決するための手段】本発明の請求項1に係る音声合成装置は、音声波形が予め記憶された波形記憶手段と、該波形記憶手段の音声波形に含まれる第1の音声素片に対して付与された第1のラベルと、当該第1の音声素片に隣接する音声素片に対して付与された第2のラベルと、当該第1の音声素片の記憶場所情報とを含む音素データが予め記憶される音素データ記憶手段と、音声合成指令としての入力音素記号列と該入力音素記号列に隣接する隣接入力音素記号列とからなる組と、前記音素データに含まれる第1のラベルおよび第2のラベルの組とを照合して、当該入力音素記号列に対する音素データを特定する音素データ特定手段と、前記波形記憶手段のなかの前記特定された音素データの示す記憶場所に記憶された音声素片を抽出する音声素片抽出手段と、該音声素片抽出手段で抽出された音声素片を所定の順序で接続し出力する音声合成手段とを有する。

【0012】本発明の請求項1に係る音声合成装置では、音声波形が予め記憶された波形記憶手段が設けられ、さらに、この波形記憶手段の音声波形に含まれる第1の音声素片に対して付与された第1のラベルと、第1の音声素片に隣接する音声素片に対して付与された第2のラベルと、第1の音声素片の記憶場所情報とを含む音素データが予め記憶される音素データ記憶手段が設けられ、音声合成指令としての入力音素記号列が入力されると、音素データ特定手段により、その入力音素記号列と隣接入力音素記号列とからなる組と、音素データに含まれる第1のラベルおよび第2のラベルの組とが照合され、入力音素記号列に対する音素データが特定され、その後、音声素片抽出手段により、波形記憶手段のなかの

特定された音素データの示す記憶場所から、逐次音声素片が抽出され、音声合成手段により、適切な合成音となるような順序で、これら音声素片が接続され、合成音が出力される。

【0013】また、本発明の請求項2に係る音声合成装置は、請求項1記載の音声合成装置において、前記音素データには、前記第1の音声素片と該第1の音声素片に隣接する音声素片との間の状態に係る時間情報が付与され、該時間情報に基づいて音声素片の接続部分の状態を設定する時間設定手段を有することを特徴とする。

【0014】本発明の請求項2に係る音声合成装置では、音素データには、第1の音声素片と、該第1の音声素片との間の状態に係る時間情報が予め付与され、時間設定手段が時間情報に基づいて、音声素片の接続部分の状態を設定して、音声上の人の息継ぎ等の特徴が再現可能となる。

【0015】また、本発明の請求項3に係る音声合成装置は、請求項2記載の音声合成装置において、前記時間設定手段は、前記音素データに、前記第1の音声素片と該第1の音声素片の前後に隣接する各音声素片との間の無音部分の時間情報がそれぞれ付与されているとき、前記第1の音声素片の後の無音部分の時間情報と、当該第1の音声素片の後に接続される音声素片の前の無音部分の時間情報の少なくとも一方に対し重み付けを行って、当該各音声素片を接続する部分に設ける無音時間を算出することを特徴とする。

【0016】本発明の請求項3に係る音声合成装置では、音素データには、予め第1の音声素片と該第1の音声素片の前後に隣接する各音声素片との間の無音部分の時間情報がそれぞれ付与され、時間設定手段により、第1の音声素片の後の無音部分の時間情報と、当該第1の音声素片の後に接続される音声素片の前の無音部分の時間情報の少なくとも一方に対して重み付けが行われて、各音声素片を接続する部分に設ける無音時間が算出され、どちらか一方の音声素片の特徴に偏った無音時間が設定されることが回避され、たとえ実音声上で隣接していない音声素片どうしを接続する際にあっても、自然な合成音が得られる。

【0017】また、本発明の請求項4に係る音声合成装置は、請求項1ないし請求項3のいずれかに記載の音声合成装置において、前記入力音素記号列に対して特定された音素データが、当該入力音素記号列の前又は後ろに隣接する入力音素記号列に対して特定されることを規制する規制手段を有することを特徴とする。

【0018】本発明の請求項4に係る音声合成装置では、入力音素記号列に対して特定された音素データは、当該入力音素記号列の前又は後ろに隣接する入力音素記号列に対して特定されず、実音声上では連続することが極めて希な同一の音声素片が特定されて合成音が不自然となる不都合が回避される。

【0019】また、本発明の請求項5に係る音声合成装置は、請求項1ないし請求項4のいずれかに記載の音声合成装置において、前記音声合成手段により出力される音声素片の出力タイミングに係る情報を生成するタイミング情報生成手段を具備することを特徴とする。

【0020】本発明の請求項5に係る音声合成装置では、音声合成手段により出力される音声素片の出力タイミングに係る情報が、タイミング情報生成手段によって生成され、この情報によって、顔画像における口の形状を変化させるタイミングを決定することができるようにしている。

【0021】また、本発明の請求項6に係る音声合成装置は、請求項5記載の音声合成装置において、前記タイミング情報生成手段は、前記音声合成手段により出力される音声素片に当該音声素片よりも短い音声が含まれるときには、当該音声との同期をとるための情報を前記出力タイミングに係る情報に含ませることを特徴とする。

【0022】本発明の請求項6に係る音声合成装置では、音声合成手段により出力される音声素片に、例えば母音や子音などの音素や母音および子音からなる音節などの、当該音声素片よりも短い音声が含まれるときには、タイミング情報生成手段により、当該音声との同期をとるための情報を出力タイミングに係る情報に含ませて、顔画像における口の形状を音声素片よりも短い音声に同期させて変化させられるようにしている。

#### 【0023】

【発明の実施の形態】以下、本発明に係る音声合成装置の実施の形態を図1ないし図11を参照して説明する。尚、以下の説明では、日本語テキストからの音声合成を例にするが、音素分解処理部および韻律処理部を変えれば、他の言語で入力されたテキストからの音声合成が可能である。

【0024】図1は、本発明に係る音声合成装置の第1の実施の形態を示す図である。

【0025】本実施の形態の音声合成装置には、キーボードや、予め用意されたテキストデータを保存するハードディスク装置等の外部記憶装置が接続される。そして、この音声合成装置には、所定の処理を実行する演算部、および処理命令の記憶やデータの一時保存が可能な主記憶部とが設けられており、この主記憶部や、外部記憶装置に記憶された命令を逐次演算部に読み込ませ、実行させることにより以下の音声合成のための処理が行われる。また、以下に説明する各部間のデータの転送については、主記憶部に設けられた所定の保存領域、或いは命令によって逐次設定される作業領域を介して行われる。

【0026】尚、必要に応じてマウス、ディスプレイ装置等の入出力装置を接続してもよい。また、後述する実音声波形記憶部17と音素データ記憶部とを、複数の端末との通信が可能なコンピュータ等に配備して、ネット

ワーク型の音声合成装置を構成することも可能である。

【0027】本図に示すテキスト入力部1は、キーボードや外部記憶装置等から、入力バッファを介して、日本語テキストデータを入力し、そのデータを音素分解処理部3、および韻律処理部5へと出力する。音素分解処理部3は、テキスト入力部1から転送された日本語テキストデータに言語処理を行うとともに、音声素片（この音声素片には、単音素および、単音素が複合された音素列が含まれる）の種類を表し、予め後述する音素データ記憶部19に登録されている音素記号あるいは音素記号列に変換する。

【0028】以下の説明では、入力されたテキストデータを「入力テキスト」といい、この変換により得られた各音素記号列（単一の音素記号を含む）を「入力ブロック」ないしは「入力音素記号列」という。例えば、入力テキスト「おはようございます」は、この時点で8個の入力ブロック／o／、／h a／、／y o／、／u／、／g o／、／z a i／、／m a／、／s u／からなる入力ブロック列に変換される。

【0029】そして、音素分解処理部3は、各入力ブロックを韻律処理部5および音素データ候補選定部7へと送出する。韻律処理部5は、所定のアクセントパターン表や再生時間長規則を基にした韻律処理を実行することにより、音素分解処理部3から入力された各入力ブロックに対して、テキスト入力部1から入力された入力テキストを解析しながら、平均基本周波数 $f$ と再生時間長 $t$ を決定する。そして、韻律処理部5は、各入力ブロックに平均基本周波数 $f$ と再生時間長 $t$ とを対応づけて、音素データ候補選定部7へと送出する。

【0030】音素データ候補選定部7は、入力された入力ブロック、平均基本周波数 $f$ および再生時間長 $t$ と、後述する音素データ記憶部19に記憶されたデータベースに含まれるデータとを照合して、各入力ブロックに対する音素データの候補を選定する。そして、候補を表すフラグを音素データに対応づけ、後述する音素データ特定部9での処理において、候補とそうでないものの区別を可能とする。

【0031】音素データ特定部9は、前記候補を表すフラグに基づいて、各入力ブロックについての候補のなかから、合成音生成のために最適な1つの音素データを最終的に特定するまでの処理を行う。これ以降、特定された音素データを特定音素データという。この処理により、各入力ブロックに対する、合成音生成のための音声素片が抽出可能となる。そして、特定音素データを後述する音声素片抽出部11および無音時間計算部13へと出力する。

【0032】音声素片抽出部11は、特定音素データの内容を解析して、合成音を生成するための音声素片を実音声波形記憶部17から抽出する。そして音声素片を音声合成部15へと出力する。

【0033】無音時間計算部13は、音素データ特定部9から入力された特定音素データの内容を解析し、各入力ブロックに対する音声素片を接続する際の、接続部分に設ける無音時間を計算する。そして、無音時間計算部13は、無音時間を示す無音時間データを音声合成部15に送出する。

【0034】音声合成部15は、無音時間計算部13から入力された無音時間データに基づいて、無音時間を音声素片波形データの接続部に設定して、音声素片抽出部11から入力した各波形データと各無音時間を順次接続することで、1つの合成音データを生成し、これを増幅して出力する。

【0035】実音声波形記憶部17は、図1に示すように、合成音生成の際に音声素片抽出部11により抽出される音声素片を含む波形データがデータベースとして蓄積記憶された部分である。これら波形データは、例えば、小説等の所定の日本語長文を人間に読ませ、これを録音して適当な長さに分割したものを符号化したものであり、様々な録音時間の波形データが含まれている。また、同じ音声からのデータであっても、録音時の声の状態や、読まれた文章の内容等により、その波形は異なったものになる。

【0036】この実音声波形記憶部17は、各データ毎に、2つの保存領域を有し、保存領域1「音声素片データ名」には、検索用の名前が保存されている。図1の例では、音声「おはようございます」に相当する名前「ohayougouzaimasu001」が保存されている。また、保存領域2「波形データ」には、音声「おはようございます」に相当する波形データが保存されている。

【0037】そして、本図に示す音素データ記憶部19は、保存領域1から保存領域9までの9つの保存領域を有する音素データのデータベースである。この音素データ記憶部19は、音声素片に1対1に対応している（以下、この音声素片を、各保存領域の説明において「当該音声素片」という）。そして、音素分解処理の際には、入力テキストに対して入力ブロックを決定（割り当て）するために、音素分解処理部3によって参照される部分であり、また、音素データ特定部9によって、最適な音声素片を特定するために照合される部分でもある。

【0038】その保存領域1「音素記号」には、入力ブロックと対比される音素記号（または音素記号列）が登録保存される。本例では、音声「よ」に相当する音素記号／yo／が保存されている。即ち、この音素記号または音素記号列が第1のラベルとしての機能を果たす。

【0039】保存領域2「所属音声素片データ名」には、当該音声素片が、実音声波形記憶部17に生成されたデータベースの中の、どのデータに所属するかを示す名前が保存される。本図の例では、実音声「おはようござい

zaimasu001」というデータの中に、音素記号／yo／に相当する音声素片が含まれていることを示している。

【0040】保存領域3「音素再生開始時間」には、当該音声素片が、「所属音声素片データ名」の名前の示す再生音声（以下、各保存領域の説明において「当該再生音声」という）の、どの時点から開始される音声に相当するものかを示すデータが保存される。このデータとしては、トータルの再生時間に対する相対的な開始時刻を示す値等を使用すればよい。

【0041】一方保存領域4「音素再生終了時間」には、当該音声素片が、当該再生音声の、どの時点までの音声に相当するものかを示すデータが保存される。この「音素再生終了時間」にも「音素再生開始時間」と同様に、トータルの再生時間に対する相対値等を使用すればよい。従って、「所属音声素片データ名」、「音素再生開始時間」および「音素再生終了時間」により当該音声素片の記憶場所が特定可能となっている。

【0042】そして、保存領域5「直前の音素」には、当該再生音声のなかで、当該音声素片の直前の音声素片を示す音素記号（または音素記号列）が登録保存され、一方保存領域6「直後の音素」には、当該音声素片の直後に再生される音声素片を示す音素記号（または音素記号列）が登録保存される。即ち、これら保存領域「直前の音素」または「直後の音素」に登録された音素記号（または音素記号列）が第2のラベルとして機能する。

【0043】保存領域7「直前の無音部分長」には、当該音声素片とその直前の音声素片との間の無音部分の長さを示すデータが保存され、一方保存領域8「直後の無音部分長」には、当該音声素片とその直後の音声素片との間の無音部分の長さを示すデータが保存される。これら「直前の無音部分長」および「直後の無音部分長」には、実際の時間情報を保存しても良いし、また、当該音声素片の再生時間等に対する相対値を保存してもよい。尚、これら無音部分は、ポーズともいわれる。そして、音素データ記憶部19の保存領域9「平均基本周波数」には、当該音声素片の平均基本周波数を示す値が保存される。

【0044】次に、図2に示すフローチャートを参照して、本発明に係る音声合成装置の動作を実例をまじえて説明する。本図に示す処理は、入力テキストが音声合成装置の入力バッファ等に蓄えられている状態から、音声合成の指令等によって開始するものである。尚、以下の実施の形態の説明において、データベースの保存領域名を鍵かっこで囲んだ表現を使用する場合は、データベースの保存領域に保存されたデータそのものを示すものとする。

【0045】図2においては、先ず、ステップS1で、例えば「おはようございます」等の入力テキストが、入力バッファ等からテキスト入力部1へと入力される。入

力された文字列は、ステップS3において、音素分解処理部3へと出力され、音素分解処理部3では、音素データ記憶部19に保存された「音声」を参照することにより、入力テキストを複数の入力ブロックに変換する（この変換処理を音素分解という）。

【0046】上記例示した入力テキスト「おはようございます」は、この時点で8個の入力ブロック／o／、／ha／、／yo／、／u／、／go／、／zai／、／ma／、／su／からなる入力ブロック列に変換される。尚、便宜上、得られた入力ブロックの数をn個とし、各入力ブロックをそれぞれ、入力テキストの順に合わせて、第1入力ブロックc1、第2入力ブロックc2、…、第n入力ブロックcnというものとする。

【0047】韻律処理部5は、次のステップS5において、各入力ブロックに対して、所定のアクセントパターン表や再生時間長規則に基づいて、音声再生時の音程を決定する平均基本周波数fと音声の継続時間を決定する再生時間長tとを割当てる。ここで、i番目の入力ブロックに割当てられた平均基本周波数、音素再生長をそれぞれfi、tiというものとする。尚、入力ブロックが単一の音素記号の複合されたものである場合は、その音素記号列の各音素記号に割当てられた基本周波数を平均化したものが平均基本周波数fとなる。

【0048】そして、韻律処理部5は、平均基本周波数fと再生時間長tとを入力ブロックに対応づけて、主記憶部の所定の保存領域に保存する。尚、入力ブロック、平均基本周波数および再生時間長を総称して入力データという。

【0049】そして、図2のステップS7においては、音素データ候補選定部7は、第1入力データ、即ち、第1入力ブロック、平均基本周波数f1および再生時間長t1を読み込み、所定の作業領域に保存する。ステップS9においては、音素データ候補選定部7は、各入力データを音素データ記憶部19に記憶されたデータに対比させて、音素データの候補を選定する。

【0050】ここで、図3に示すフローチャートを参照して、音素データ候補選定部7にて各入力ブロックについて実行される、音素データの候補選定処理について説明する。この候補選定処理は、最終的に、入力ブロックに対して1つの音素データを特定するための前段階として、実行されるものであり、先ず、「音声記号」が入力ブロックに一致し、しかも「平均基本周波数」および音声素片の再生時間が、入力ブロックに割り当てられたものに対して、所定の誤差範囲に含まれる音素データが候補として選定される。選定にあたっては、候補を表すフラグを音素データに対応づけ、後の処理のために、このフラグをオン／オフいずれかの状態に設定する。

【0051】図3におけるステップS41では、音素データ候補選定部7は、入力ブロックciを検索パラメータとして音素データ記憶部19に記憶された全音素デー

タの保存領域1を検索し、入力ブロックと「音声記号」が一致する音素データを候補として選定する。音素データ記憶部19には、通常「音素記号」が同一である多数の音素データが保存されているため、通常は、この時点での候補数は比較的多い。

【0052】音素データ候補選定部7は、前述のフラグの状態をもとに、すべての候補について、順次に以下の処理を行う。

【0053】先ずステップS43において、第1の候補（音素データ）を読み込んで主記憶部の所定の作業領域に保存する。そして、ステップS45においては、当該入力ブロック（ci）に割当てられた平均基本周波数fiと、当該音素データの「平均基本周波数」との差Δf（偏差）が所定の周波数F以上であるか否かを判定する。このステップS45で、この偏差Δfが値F以上であると判定される（YES）と判定されると、ステップS47において、この音素データを候補から除外する処理（以下、候補除外処理という）が行われる。即ち、「平均基本周波数」が所定の誤差範囲にない音素データは、ここで候補から除外される。

【0054】一方、ステップS45で、偏差Δfが周波数Fより小さい（NO）と判定されると、続くステップS49においては、当該対象となっている候補の「音素再生開始時間」と「音素再生終了時間」との時間差、即ち音声素片の再生時間長と、入力ブロックciに割当てられた再生時間長tiとの偏差Δtが所定の時間値T以上であるか否かを判定する。このステップS49において、偏差Δtが値T以上であると判定されると、ステップS47において、候補除外処理が行われる。即ち、再生時間長が所定の誤差範囲にない音素データは、ここで候補から除外される。

【0055】一方、ステップS49において、偏差Δtが値Tより小さい（NO）と判定された場合は、次にステップS51にて、その候補（音素データ）が当該入力ブロックciの直前の入力ブロックci-1について特定された音素データであるか否かが判定され、直前の入力ブロックci-1について特定された音素データである（YES）と判定されたときは、制御がステップS47に移行して、その候補について候補除外処理が行われる。即ち、実音声では、同一音が連続することが極めて希であるため、音声合成装置がこのような誤った選定を行い、合成音が不自然となるのを回避する効果がある。

【0056】ステップS51にて、直前の入力ブロックci-1について特定された音素データでない（NO）と判定されたときは、その音素データは候補として残されることとなる。そして、ステップS51にてNOと判定された後、あるいは、ステップS47における候補除外処理の終了後は、音素データ候補選定部7は、ステップS53において、当該候補が最後の候補か否かを判定する。ここで、最後の候補でない（NO）と判定される

と、音素データ候補選定部7は、ステップS55に制御を移行させ、次の候補（音素データ）の内容を読み込む。そして、音素データ候補選定部7は、ステップS45からステップS53までの一連の処理を、ステップS53において、最後の候補である（YES）と判定されるまで順次実行する。

【0057】音素データ候補選定部7が上記処理を行うことにより、後述する波形の抽出時において、同一の音素記号で、しかも音声の平均基本周波数および再生時間長の近い音声素片の抽出が可能となる。

【0058】そして、図2に示すステップS11においては、例えば、図1に示したような1つの入力ブロック「yo」に対して、ある程度絞り込まれた候補のなかから最終的な1つの音素データを特定する処理を音素データ特定部9が行う。特にこの実施の形態では、入力ブロックの前後の関係を、波形データ記録時の音声素片の前後関係に対比させることにより、隣接音声環境の最適な音声素片を特定することを可能としている。

【0059】図4に示すフローチャートは、第1入力ブロックに対して、音素データ特定部9が実行する処理を説明するためのものであり、本図のステップS61においては、第2入力ブロックを主記憶部の所定作業領域に読み込む。そして、現在残っている候補のなかの最初の候補（音素データ）につき、その「直後の音素」を読み込む（ステップS63）。そして、ステップS61にて読み込んだ入力ブロックc2とステップS63にて読み込んだ「直後の音素」とが一致するか否かを、ステップS65にて判定し、ここで一致しない（NO）と判定された場合は、音素データ特定部9は、ステップS67において候補除外処理を行う。

【0060】一方、ステップS65にて、一致する（YES）と判定されたときは、候補除外処理は行われない。ステップS65にて、一致する（YES）と判定された後、あるいは、ステップS67における候補除外処理の終了後は、音素データ特定部9は、ステップS69において、当該候補が最後の候補か否かを判定する。ここで、最後の候補でない（NO）と判定されると、ステップS71に進んで、次の候補の「直後の音素」を読み込み、そして、音素データ特定部9は、ステップS65からステップS69までの一連の処理を、ステップS69において、最後の候補である（YES）であると判定されるまで順次実行する。

【0061】このようにして、第1入力ブロックについては、当該候補の「直後の音素」が第2入力ブロックと一致するものだけが、候補として残されることとなる。

【0062】図5に示すフローチャートは、第1入力ブロックおよび最後の入力ブロックを除く、任意の第i入力ブロックに対して、音素データ特定部9が実行する処理を説明するためのものである。

【0063】音素データ特定部9は、ステップS81に

において、当該入力ブロックciの直前の入力ブロックおよび直後の入力ブロック、即ち、入力ブロックci-1および入力ブロックci+1を主記憶部に読み込み、その後、ステップS83において、最初の候補の「直前の音素」および「直後の音素」を主記憶部に読み込む。そして、ステップS85においては、当該候補（音素データ）が以下のどの条件を満たすかにより候補の分類を行う。

【0064】まず、当該候補の「直前の音素」と入力ブロックci-1とが一致し、かつ、「直後の音素」と入力ブロックci+1が一致する候補に対しては、優先して候補に残される可能性（優先度）が一番高いことを示すデータP1が対応づけられる。図1に示す例では、入力ブロック「yo」に対して、「直前の音素」が「ha」であり、かつ、「直後の音素」が「u」であり、この条件を満たすため、この音素データには、データP1が対応づけられる。

【0065】そして、「直前の音素」と入力ブロックci-1が一致し、かつ、「直後の音素」と入力ブロックci+1が一致しない候補に対しては、データP1よりも優先度が低いことを示すデータP2が対応づけられる。

【0066】そして、さらに「直前の音素」と入力ブロックci-1とが一致せず、かつ、「直後の音素」と入力ブロックci+1が一致する候補に対しては、データP2よりもさらに優先度が低いことを示すデータP3が対応づけられる。

【0067】最後に、「直前の音素」と入力ブロックci-1とが一致せず、かつ、「直後の音素」と入力ブロックci+1とが一致しない候補に対しては、候補として不適当であるため、いかなるデータの対応づけも行わない。

【0068】上記の分類処理の終了後、音素データ特定部9は、ステップS87において、当該候補が最後の候補か否かを判定する。ここで、最後の候補でない（NO）と判定されると、ステップS89に制御を移行させ、次の候補の「直前の音素」と「直後の音素」とを読み込み、その後、ステップS85へと制御を移行させる。これ以降、音素データ特定部9は、ステップS85とステップ87の処理を、ステップS87にて、最後の候補である（YES）と判定されるまで繰返し行う。

【0069】ステップS87において最後の候補であると判定されると、次のステップS91においては、上記優先度について分類された候補を次のような処理によって絞り込む。

【0070】音素データ特定部9は、まず、ステップS91にて、データP1が対応づけられた候補があるか否かを判定し、データP1が対応づけられた候補がある

（YES）場合は、ステップS93において、その他の候補に対して候補除外処理を行う。一方、ステップS9

1において、データP1が対応づけられた候補がない（NO）と判定された場合は、次のステップS95において、データP2が対応づけられた候補があるか否かを判定する。データP2が対応づけられた候補がある（YES）と判定された場合は、ステップS93において、その他の候補に対して候補除外処理を行う。一方、ステップS95にて、データP2が対応づけられた候補がない（NO）と判定された場合は、次のステップS97にて、データP3が対応づけられた候補があるか否かを判定する。ここで、データP3が対応づけられた候補がある（YES）と判定された場合は、ステップS93において、その他の候補に対して候補除外処理を行う。そして、ステップS97において、データP3が対応づけられた候補がない（NO）と判定された場合、即ち、全ての候補に対し、データP1、P2およびP3のいずれのデータの対応づけもなされていない場合は、音素データ特定部9は、ステップS99において、全ての候補について候補除外処理を行う。そして、ステップS99における処理、あるいは、ステップS93における処理の終了により、当該入力ブロックciについての処理を終える。

【0071】図6に示すフローチャートは、最後の入力ブロック、即ち入力ブロックcnに対して、音素データ特定部9が実行する処理を説明するための図であり、本図におけるステップS111においては、入力ブロックcn-1を主記憶部に読み込む。そして、現在残っている候補のなかの最初の候補（音素データ）の「直前の音素」を読み込む（ステップS113）。そして、ステップS111にて読み込んだ入力ブロックcn-1とステップS113にて読み込んだ「直前の音素」とが一致するか否かをステップS115にて判定し、ここで一致しない（NO）と判定された場合は、音素データ特定部9は、ステップS117において当該候補について候補除外処理を行う。

【0072】一方、ステップS115にて、一致する（YES）と判定された場合は、その候補は、候補としての適格性があるので、候補除外処理は行われない。ステップS115にてYESと判定された後、あるいは、ステップS117における候補除外処理の終了後は、音素データ特定部9は、ステップS119において、当該候補が最後の候補か否かを判定する。ここで、最後の候補でない（NO）と判定されると、ステップS121に移行して、次の候補の「直前の音素」を読み込み、そして、音素データ特定部9は、ステップS115からステップS119までの一連の処理を、ステップS119において、最後の候補である（YES）であると判定されるまで順次実行する。

【0073】従って、図4ないし図6を参照しながら説明したように、この実施の形態の音声合成装置では、音素データ特定部9は、互いに隣接する入力ブロックを、

音素データの「直前の音素」あるいは「直後の音素」に対比させながら、候補を選定するため、隣接音声環境が適切な音声素片を抽出することができ、音声素片の接続部分のノイズを低減して、録音された実音声の特徴を備え、連続性が良好な自然な合成音を得られるという効果がある。

【0074】次に、音素データ特定部9は、上記のように選定された候補を最終的に1つの音素データにまで特定する処理を行う。

10 【0075】図7に示すフローチャートにおいて、音素データ特定部9は、先ずステップS131で、現在残っている候補の数を算出する。次のステップS133では、音素データ特定部9は、ステップS131での候補数算出の結果によって、候補が存在するか否かを判定する。このステップにおいて、候補が存在しない（NO）と判定された場合、即ち、候補がすべて候補除外処理により除外されていた場合は、入力ブロック、平均基本周波数および再生時間長による候補の選定処理（図2のステップS9）が終了した時点で候補であったものを、再び候補とする処理を行う（ステップS135）。

20 【0076】一方、ステップS133にて、候補が存在すると判定された場合（YES）あるいは、ステップS135での処理の終了後は、音素データ特定部9は、その候補のなかで、入力ブロックに対して割り当てられた平均基本周波数fと当該候補の「平均基本周波数」との偏差 $\Delta f$ が最小となる音素データを最終的に特定する（ステップS137）。

30 【0077】以上のような処理過程を経て、図2に示すステップS11では、各入力ブロックに対して、最終的に1つの音素データが特定される。

【0078】そして、図2において、ステップS13では、現在処理を行っている入力ブロックが、最後の入力ブロック（第n入力ブロック）であるか否かが判定され、ここで最後の入力ブロックでない（NO）と判定されると、ステップS15において、次の入力データを主記憶部に読み込む。そして、制御をステップS9に移行させ、入力ブロックに対しての候補選定を行う。これ以降は、ステップS13において、最後の入力ブロックである（YES）と判定されるまで、上記処理を順次実行する。

40 【0079】そして、ステップS13にて、最後の入力ブロックであると判定されると、次に音声素片抽出部11により、音声素片の抽出が行われる（ステップS17）。

【0080】次に、音声素片の抽出過程を図8に示すフローチャートを参照して説明する。このフローチャートは、1つの入力ブロックについての音声素片の抽出過程を示すものである。

50 【0081】本図におけるステップS141において、入力ブロックに対応した特定音素データの「所属音声素



片データ名」、「音素再生開始時間」および「音素再生終了時間」を読み込む。次に、ステップS143においては、図1に示すように、実音声波形記憶部17に記憶されたデータベースのなかの「所属音声素片データ名」(ohayougozaimasu001)が示す波形データの範囲において、特定音素データの「音素再生開始時間」の示す時点(3000)から「音素再生終了時間」の示す時点(4000)までの部分的な波形データ(音声素片)を所定の作業領域に読み込み、このデータを合成音生成のための所定の保存領域に保存する。そして、音声素片抽出部11は、以上の処理をすべての入力ブロックについて行う。尚、各入力ブロックに対して抽出され保存された音声素片を、音声素片SU1, SU2, ..., SUnとする。

【0082】次に無音時間計算部13は、図2に示すように、全ての音声素片の抽出が終了すると、ステップS19において、音声素片を接続する際の接続部分に設ける無音時間を計算する。即ち、この実施の形態は、この無音時間の設定により、音声上の人の息継ぎ等の特徴を再現させようとするものである。

【0083】次に、この計算方法を図9に示すフローチャートを参照して説明する。このフローチャートは、1つの音声素片と、これと隣接する音声素片との間の接続部分に設けられる無音時間の計算方法を示すものである。

【0084】無音時間計算部13は、ステップS151において、最後の入力ブロックを除いて、入力ブロックに対応する特定音素データの「直後の無音部分長」を読み込む(これをデータAとする)。次に、ステップS153において、当該入力ブロックciの直後の入力ブロックci+1に対応する特定音素データの「直前の無音部分長」を読み込む(これをデータBとする)。

【0085】次に、ステップS155では、条件として、データA=0であり、かつ、データB=0であるか否かが判定される。この条件を満たす(YES)ときは、無音時間計算部13は、音声素片SUiと音声素片SUi+1の接続部分の無音時間は0(ms)とする(ステップS159)。一方、ステップS155の条件を満たさない(NO)とき、即ち、データAおよびデータBのいずれか一方が0でないときは、ステップS157において、無音時間=(A+B)/2(ms)に設定する。そして、これら無音時間をデータ(無音時間データ)として所定の保存領域に保存する。

【0086】例えば、実音声「おはようございます(ohayougozaimasu)」に対する波形データが実音声波形記憶部17に記憶され、音素記号/u/に対する「直後の無音部分長」が150msである場合にあって、入力テキスト「おはよう」に続いて入力された「いいんきですね」の「う」に対して、前記の音素記号/u/が特定されると、前述の無音時間150msに

については、続く入力テキスト「いいんきですね」に対しては短すぎてしまうが、入力テキスト「い」に対する特定音素データの「直前の無音部分長」が例えば500msとすれば、上記の無音時間の設定方法により計算される平均、(150ms+500ms)/2=325msを設定することで、上述の不都合を回避することができる。尚、この実施の形態では、平均を計算する方法を採用したが、例えば、前の音声素片に対する「直後の無音部分長」の方を重視する等の、状況に応じた重み付け演算を行ってもよい。

【0087】以上の説明から明らかなように、この実施の形態では、音声上の人の息継ぎ等の特徴が再現可能となる効果が得られる。加えて、無音時間を介して接続される、各音声素片を抽出するための各音素データの「直前の無音部分長」、「直後の無音部分長」に重み付けを行うことにより、偏った無音時間が設定されるのを回避して、実音声上では前後しない音声素片どうしを接続する際にあっても、自然な合成音が得られる。

【0088】そして、無音時間計算部13は、最後の音声素片SUnを除く全ての音声素片に対して上記処理を行うことで、音声素片の接続部分に設ける無音時間を決定し、これらデータを主記憶部の所定保存領域に記憶する。

【0089】最後に、音声合成部15は、図2のステップS21において、主記憶部の所定領域に保存された各音声素片と、接続部分に設ける無音時間を示す無音時間データとを読み出し、これらを順次接続して1つの合成音の波形データとする。そして、ステップS23において、合成された波形データをA/D変換してアナログ信号に変換し、そのアナログ信号を増幅して合成音として出力する。

【0090】従って、本発明に係る音声合成装置の第1の実施の形態によれば、音声波形が予め記憶された実音声波形記憶部17が設けられ、さらに、音素データ記憶部19には、実音声波形記憶部17の音声波形に含まれる音声素片(第1の音声素片)に対して付与され「保存領域1」に保存された音素記号(第1の音素記号)と、その第1の音声素片に隣接する音声素片に対して付与され「保存領域5」あるいは「保存領域6」に保存された「直前の音素」あるいは「直後の音素」、即ち第2の音素記号と、「保存領域2」の「所属音声素片データ名」、「保存領域3」の「音素再生開始時間」並びに「保存領域4」の「音素再生終了時間」からなる第1の音声素片の記憶場所情報とを含む音素データが予め記憶され、音声合成指令として入力音素記号列(入力ブロック)が入力されると、音素データ特定部9により、その入力音素記号列とこれに隣接する入力音素記号列とからなる組と、音素データ記憶部19における第1の音素記号および第2の音素記号の組とが照合され、入力音素記号列に対する音素データが特定され、その後、音声素片

抽出部11により、実音声波形記憶部17のなかの特定された音素データの示す記憶場所から、逐次音声素片が抽出され、音声合成部15により、所定の順序でこれら音声素片が接続され、合成音が出力されるため、音声素片の接続部分のノイズを低減するとともに、録音された実音声の特徴を備えた高品質の合成音出力が可能となる。

【0091】次に、本発明に係る音声合成装置の第2の実施の形態について説明する。

【0092】この実施の形態は、上記第1の実施の形態と同様に高品質の合成音を得るとともに、合成音を構成する音声素片の出力タイミングに係る情報を含む文字列L（以下、リップシンク信号列Lという）を生成することを特徴とする。即ち、この実施の形態では、このリップシンク信号列Lを、例えばCGを制作するための画像処理装置などで解読して、顔画像における口の形状を変化させるタイミングを決定することができるようにしている。

【0093】具体的には、第2の実施の形態の音声合成装置には、タイミング情報生成部21が設けられ、このタイミング情報生成部21においてリップシンク信号列Lが生成される。

【0094】図10は、このリップシンク信号列Lを生成するための処理を示すフローチャートであり、このフローチャートを参照しながら、入力テキスト「こんにちは」に係る処理を説明していく。

【0095】先ず、上記第1の実施の形態と同様に、入力テキスト「こんにちは」は、図1に示す音素分解処理部3によって、4個の入力ブロック／k o／、／n／、／n i／、／t i w a／に変換される。そして、音素データ候補選定部7および音素データ特定部9によって、音素データ記憶部19の中から、各入力ブロックに対して音素データが特定され、無音時間計算部13によって、入力ブロック／k o／と／n／のそれぞれに対応する音声素片の接続部分の無音時間500msが決定される。

【0096】タイミング情報生成部21は、図10のステップS161において、各入力ブロックに対して特定された音素データの「音素再生開始時間」と「音素再生終了時間」との差を再生時間長として演算し、入力ブロックと再生時間長とからなる組を再生される順序でメモリなどに保存することにより、表1に示すような「合成音再生データ」を構築する。このとき、タイミング情報生成部21は、無音時間500msを1つの音声素片の再生時間長とみなし、これに「無音」を表す／q／を入力ブロックとして対応づける。

【0097】

【表1】

合成音再生データ

i	1	2	3	4	5
ci	/ko/	/q/	/n/	/ni/	/tiwa/
ti	2500	500	1000	2000	4000

表1においてiは、再生の順序を表す変数であり、この実施の形態では、変換された4個の入力ブロックと前述の／q／とを合わせて、その最大値n=5となっている。また、ciは各入力ブロックを示す変数であり、そして、tiは、各入力ブロックに対して抽出された音声素片の再生時間長を示す変数であって、t2としての無音時間500msが含まれている。

【0098】図10に戻り、タイミング情報生成部21は、ステップS163において、リップシンク信号列Lの内容をクリアして、何も文字が含まれていない状態

（図中では、記号” ”で示す）とする。また、先頭の音声素片の出力開始時からの累積経過時間を表す変数Tを0（零）にリセットするとともに再生順序を表す変数iを1とする。即ち、これらの処理が初期化動作として実行される。

【0099】そして、タイミング情報生成部21は、ステップS165において、現在のリップシンク信号列Lに対して、入力ブロックciを右側から結合する（以下、単に「右結合する」という）。尚、このステップS165の処理が行われる前においては、L=” ”であるから、このステップS165の処理後において、L=”k o”となる。尚、結合時には、入力ブロックの前後の記号／は取除かれる。

【0100】続くステップS167では、i=1であるか否かが判定される。ここで、i=1（YES）と判定された場合は、ステップS169にて、タイミング情報生成部21は、リップシンク信号列Lに文字”0（零）”を右結合する。

【0101】その後ステップS175へと進み、ここで、処理中の当該入力ブロックciが最後の入力ブロックである（変数i=n）か否かが判定され、ここでNOと判定されたときは、ステップS177にて、変数iが1繰上げられ（i=i+1）、再びステップS165の処理が行われる。

【0102】変数iが繰上げられてi=2となったときは、ステップS165にて入力ブロック／q／が右結合され、リップシンク信号列L=”k o o q”となる。続くステップS167では、i≠1（NO）と判定されるため、ステップS171へと進む。

【0103】このステップS171では、タイミング情報生成部21は、累積経過時間を表す変数Tに対して、現在の変数iの1つ前のi-1に対応する再生時間長t（i-1）を加算処理する。即ち、t1=2500が加算され、変数T=2500となる。続いて、ステップS173では、タイミング情報生成部21は、変数Tの内

容を文字列に変換する処理を行い、変換後の文字列 c h (T) をリップシンク信号列 L に対して右結合する。従って、 $i = 2$  のときは、ステップ S 173 の処理後のリップシンク信号列 L は  $L = "k o o q 2 5 0 0"$  となる。

【0104】このようにして、タイミング情報生成部 21 は、リップシンク信号列 L を  $"k o o q 2 5 0 0" \rightarrow "k o o q 2 5 0 0 n 3 0 0 0" \rightarrow "k o o q 2 5 0 0 n 3 0 0 0 n i 4 0 0 0"$  と生成していき、最終的にリップシンク信号列  $L = "k o o q 2 5 0 0 n 3 0 0 0 n i 4 0 0 0 t i w a 6 0 0 0"$  を完成させる。そして、ステップ S 175 において当該入力ブロック c i が最後の入力ブロックである ( $i = n$ ) と判定されると、このリップシンク信号列 L の生成処理が終了する。

【0105】こうして生成されたリップシンク信号列 L に含まれる各数値は、合成音における先頭の音声素片の出力開始時間 (0) と、これを基準とした 2 番目以降の各音声素片の出力開始時間を表している。即ち、本例では、入力ブロック  $c 1 = / k o /$  に対応する音声素片「こ」の出力開始時間を基準時とすると、2500ms 後に無音状態となり、基準時から 3000ms 後に音声素片「ん」が出力され、基準時から 4000ms 後に音声素片「に」が出力され、最後に音声素片「ちわ」が、基準時から 6000ms 後に出力されることを示している。

【0106】その後、生成されたリップシンク信号列 L は、画像処理装置などにて解読され、顔画像における口の形状を変化させるタイミングが決定される。

【0107】従って、本発明に係る音声合成装置の第 2 の実施の形態によれば、タイミング情報生成部 21 を設けたことにより、音声素片の出力タイミングに係る情報を含むリップシンク信号列 L を生成することができ、このリップシンク信号列 L を利用して、音声波形を解析することなく、顔画像における口の形状を最適なタイミングで変化させることができる。

【0108】尚、本例の音声素片「ちわ」のように、音声素片が複数の音素からなる場合にあっては、その音声素片に、最初の音素、本例でいえば、 $/ t /$  に対する口の形状を対応させるか、あるいは最後の音素、本例でいえば、 $/ a /$  に対する口の形状を対応させるかの 2 通りの方法があるが、後者を選択することにより違和感の少ない顔の表情を得ることができる。

【0109】また、上記のリップシンク信号列 L における文字の並び方は、音声素片が出力されるタイミングを判読しやすいものとしたが、必ずしもこの並び方に限定されるものではなく、所定の規則により、音声素片が出力されるタイミングを解読可能とさえすればよい。

【0110】ところで、上記第 2 の実施の形態では、リップシンク信号列 L に含まれるタイミングに係る情報を、各音声素片ごとに求めたが、音声素片よりも短い音

声ごとの再生時間長を所定の規則により求め、この再生時間長を用いて、音声素片よりも短い音声との同期をとるためのタイミング情報をリップシンク信号列 L に含ませることにより、このリップシンク信号列 L を用いて、画像処理装置などにおいて、顔画像における口の形状をより滑らかに変化させることができる。

【0111】そこで、音声素片よりも短い音声との同期をとるための情報を含むリップシンク信号列 L を生成可能とした本発明の音声合成装置の第 3 の実施の形態を説明する。

【0112】図 11 は、第 3 の実施の形態の構成の一部を示すブロック図である。

【0113】この実施の形態では、リップシンク信号生成部 21 には、再生時間演算部 21a が設けられ、ここには、予め、母音および子音それぞれの再生時間長を求めるための比率  $x$  および  $y$  が設定されている。

【0114】次に、第 3 の実施の形態における音声合成装置の動作を説明する。

【0115】タイミング情報生成部 21 は、先ず、前記の表 1 に示すような合成音再生データをメモリなどから読み込み、この合成音再生データに含まれる入力ブロックの中から、適宜、分解が必要な入力ブロックを選択して、音素分解処理部 3 へと供給する。

【0116】具体的には、例えば、再生時間長  $t i$  が長い場合は、音声と画像の同期が不自然に感じられやすいため、所定の再生時間長以上のデータを選択して、音素分解処理部 3 へと供給すればよい。前述の合成音「こんにちは」を例にすると、音声素片「ちわ」の再生時間長  $t 5$  が比較的長いので、入力ブロック  $/ t i w a /$  が音素分解処理部 3 へと供給される。

【0117】音素分解処理部 3 は、供給された「分解前の入力ブロック」を音素記号に分解する。この過程において、音素分解処理部 3 は、「分解前の入力ブロック」から、母音と子音とからなる音節に相当する音節記号を抽出し、さらに、音節記号を母音あるいは子音に相当する音素記号に分解する。

【0118】そして、音素分解処理部 3 は、この分解処理によって得られた音素記号に、音節の区切を解析できるような音節区切データを付加して、タイミング情報生成部 21 に返送する。この音節区切データは、例えば、 $"12, 34"$  のようなデータに、「1 番目と 2 番目の音素記号が、母音および子音にそれぞれ相当する 1 つの音節記号を構成し、これに対して、(カンマ) で区切られた 3 番目と 4 番目の音素記号がもう 1 つの音節記号を構成する」という意味をもたせることによって生成することが可能である。

【0119】尚、音節とは、必ずしも母音と子音との組合わせに限られるものではなく、例えば、音声「ん」に相当する音素記号「n」を音節とすることもできる。本例では、「分解前の入力ブロック」 $c 5 = / t i w a /$

が、/t/、/i/、/w/、/a/の各音素記号に分解され、前述の音節区切データ”12, 34”とともにタイミング情報生成部21へと返送される。

【0120】続いて、タイミング情報生成部21の再生時間演算部21aは、音節区切データを解析し、音素分解処理部3から返送された音素記号に複数の音節記号が含まれるような場合にあっては、その各音節記号に対し当該各音節記号に対応する音節の再生時間長を与える。本例では、分解前の入力ブロックc5=/tiwa/には、/ti/および/wa/の2個の音節記号が含まれるので、これら各音節記号に対して、均等な音節の再生時間長が与えられる。即ち、音節の再生時間長=(分解前のt5)/(音節数)=4000/2=2000(ms)となる。このようにして、本実施の形態では、音節との同期をとるためのタイミング情報をリップシンク信号列Lに含ませることができるようにになっている。

【0121】そして、再生時間演算部21aは、各音節記号の中の音素記号/t/、/i/、/w/、/a/のうちの母音記号に対しては、音節の再生時間長に対する母音の再生時間長の比率として設定された比率xを与え、子音記号に対しては、比率y(ただしy=1-x)を与え、これら比率によって、音声素片に含まれる\*

合成音再生データ

i	1	2	3	4	5	6	7	8
cl	/ko/	/a/	/n/	/ni/	/t/	/l/	/w/	/a/
ti	2500	500	1000	2000	800	1200	800	1200

合成音再生データの再構築後は、上記第2の実施の形態と同様に、タイミング情報生成部21によって、合成音再生データからリップシンク信号列Lが生成される。表2の再構築後の合成音再生データからは、リップシンク信号列L=”kooq2500n3000ni4000t6000i6800w8000a8800”が生成される。そして、このリップシンク信号列Lが画像処理装置などにて解釈され、顔画像における口の形状を変化させるための、より詳細なタイミングが決定される。尚、「分解前の入力ブロック」を音素記号まで分解せずに、音節記号までの分解にとどめることも勿論可能である。

【0126】以上説明したように、この第3の実施の形態によれば、合成音として出力される音声素片に、母音あるいは子音などの音素、あるいは母音および子音からなる音節などの、音声素片よりも短い音声が含まれる場合にあっては、再生時間演算部21aにより、その音声ごとの再生時間長が演算され、タイミング情報生成部21によって、音声素片よりも短い音声との同期をとるための情報を含むリップシンク信号列Lが生成されるため、画像処理装置などにおいて、顔画像における口の形状をさらに細かく、滑らかに変化させることができる。

【0127】また、音声素片に母音および子音からなる音節が含まれるときにあっては、予め設定された、音節の再生時間長に対する母音および子音それぞれの再生時

\*母音/子音の再生時間長を演算することで、音素との同期をとるためのタイミング情報をリップシンク信号列Lに含ませることができるようにになっている。

【0122】具体的には、x=0.4とし、y=0.6とした場合に、母音および子音の再生時間長は以下のように演算される。

【0123】母音の再生時間長=音節の再生時間×比率x=2000×0.4=800(ms)

子音の再生時間長=音節の再生時間×比率y=2000×0.6=1200(ms)

タイミング情報生成部21は、演算された各再生時間長と各音素記号とを互いに対応づけて、分解前の再生時間長および入力ブロックに置換えることで、合成音再生データを再構築する。

【0124】具体的には、表2に示すような合成音再生データが再構築される。即ち、表1に示した合成音再生データにおける5番目の再生時間長および入力ブロックが、表2に示す5番目ないし8番目の再生時間長およびその入力ブロックに置換えられる。

【0125】

【表2】

間長の比率xおよびyによって、母音および子音のそれぞれの再生時間長が演算され、母音あるいは子音と同期をとるための情報がリップシンク信号列Lに含まれることとなるため、顔画像における口の形状を母音あるいは子音と同期をとって変化させることができる。また、比率x、yの設定を変えることによって、母音あるいは子音と同期をとるための情報を調整することができる。

【0128】尚、本発明に係る音声合成装置は、上記実施の形態に限るものではなく、ソフトウェアで構成された上記各処理を実行する部分を、機械読み取り可能な記録媒体に記録することも可能である。また、上記実施の形態は、音声合成の方法あるいはタイミングに係る情報の生成方法などのアルゴリズムとしても優れたものである。

【0129】そして、上記各機能を集積回路等のハードウェアで構成しても同様の効果を得ることができ、さらに、単一のコンピュータ上でのみならず、ネットワークを構成する端末やサーバマシン、あるいは、画像処理装置等に各機能を分散配備させてもよい。

【0130】

【発明の効果】以上説明したように、本発明の請求項1に係る音声合成装置によれば、音素データ特定手段により、入力音素記号列と隣接入力音素記号列とからなる組と、音素データに含まれる第1のラベルおよび第2のラ

ベルの組とが照合され、入力音素記号列に対して、隣接する音声環境の最適な音声データが特定されるため、音声素片の接続部分のノイズが低減されるとともに、実音声の特徴を備えた高品質の合成音出力が可能となる。

【0131】また、本発明の請求項2に係る音声合成装置によれば、音声素片の接続部分の状態が、音素データに付与された時間情報に基づいて設定されるため、実音声上の息継ぎ等の特徴が再現可能となる。

【0132】また、本発明の請求項3に係る音声合成装置によれば、無音部分の時間情報に対して重み付けが行われ、音声素片を接続する部分に設ける無音時間が算出されるため、どちらか一方の音声素片の特徴に偏った無音時間が設定されることが回避され、たとえ実音声上で隣接していない音声素片どうしを接続する際にあっても、自然な合成音が得られる。

【0133】また、本発明の請求項4に係る音声合成装置によれば、実音声上では連続することが極めて希な同一の音声素片の特定が規制されて、合成音が不自然になる不都合が回避される。

【0134】また、本発明の請求項5に係る音声合成装置によれば、タイミング情報生成手段により音声素片の出力タイミングに係る情報が生成されるため、音声波形を解析することなく、顔画像の口の形状を合成音に同期させて変化させることができる。

【0135】また、本発明の請求項6に係る音声合成装置によれば、タイミング情報生成手段により、音声素片よりも短い音声との同期をとるための情報が出力タイミングに係る情報に含ませられるため、口の形状をより滑らかに変化させることができる。

#### 【図面の簡単な説明】

【図1】本発明に係る音声合成装置の第1の実施の形態を示す図である。

【図2】図1に示した形態における音声合成に関する処理過程を示したフローチャートである。

【図3】図1に示した形態における入力ブロック、平均

基本周波数および再生時間長に基づく音声データの候補選定過程を示したフローチャートである。

【図4】図1に示した形態における第1入力ブロックについての候補選定過程を示したフローチャートである。

【図5】図1に示した形態における第1入力ブロックおよび最後の入力ブロックを除く入力ブロックについての候補選定過程を示したフローチャートである。

【図6】図1に示した形態における最後の入力ブロックについての候補選定過程を示したフローチャートである。

【図7】図1に示した形態における入力ブロックについての音素データの特定過程を示したフローチャートである。

【図8】図1に示した形態における音声素片の抽出過程を示したフローチャートである。

【図9】図1に示した形態における無音時間の計算過程を示したフローチャートである。

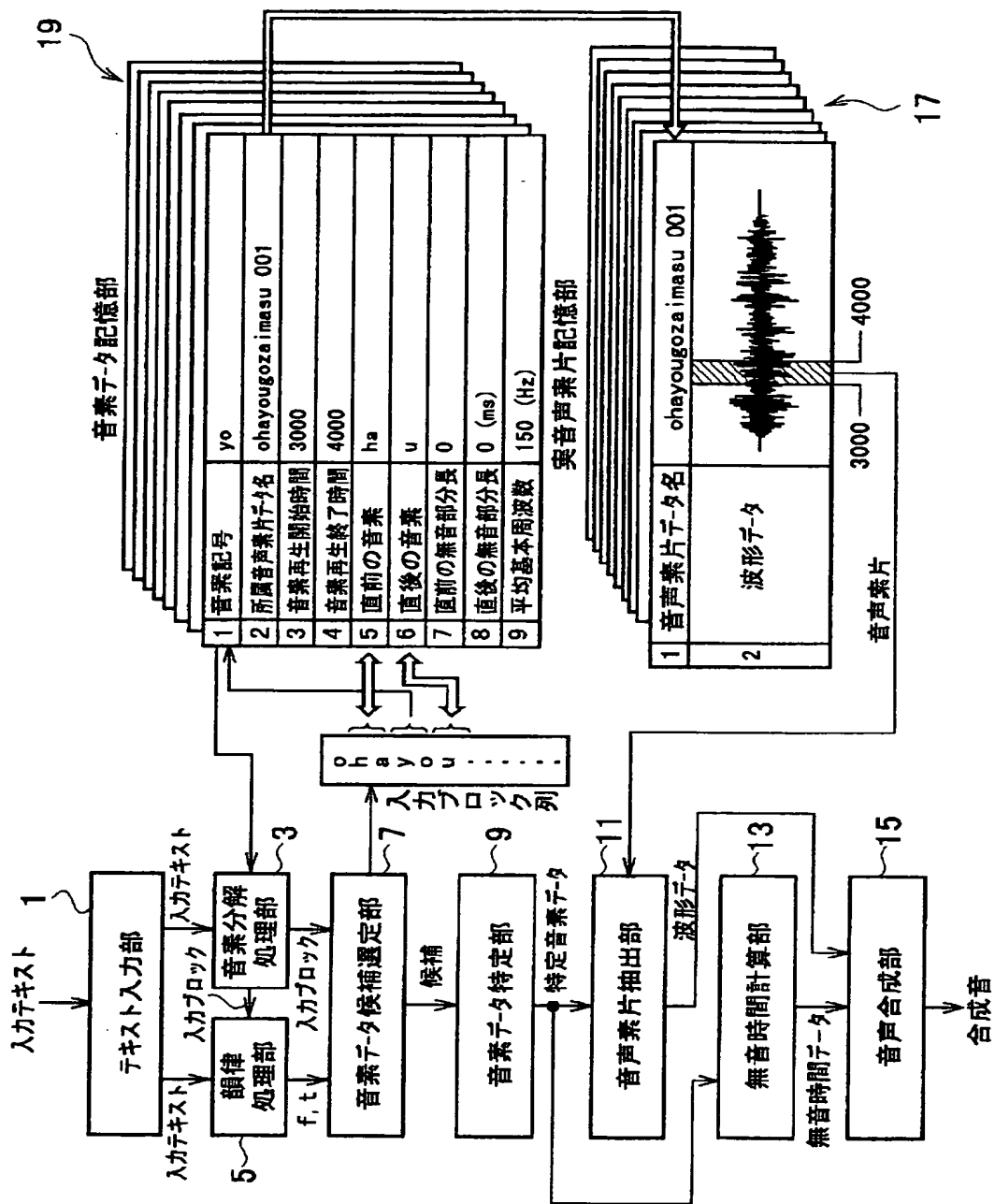
【図10】本発明に係る音声合成装置の第2の実施の形態の動作を示すフローチャートである。

【図11】本発明に係る音声合成装置の第3の実施の形態の構成の一部を示すブロック図である。

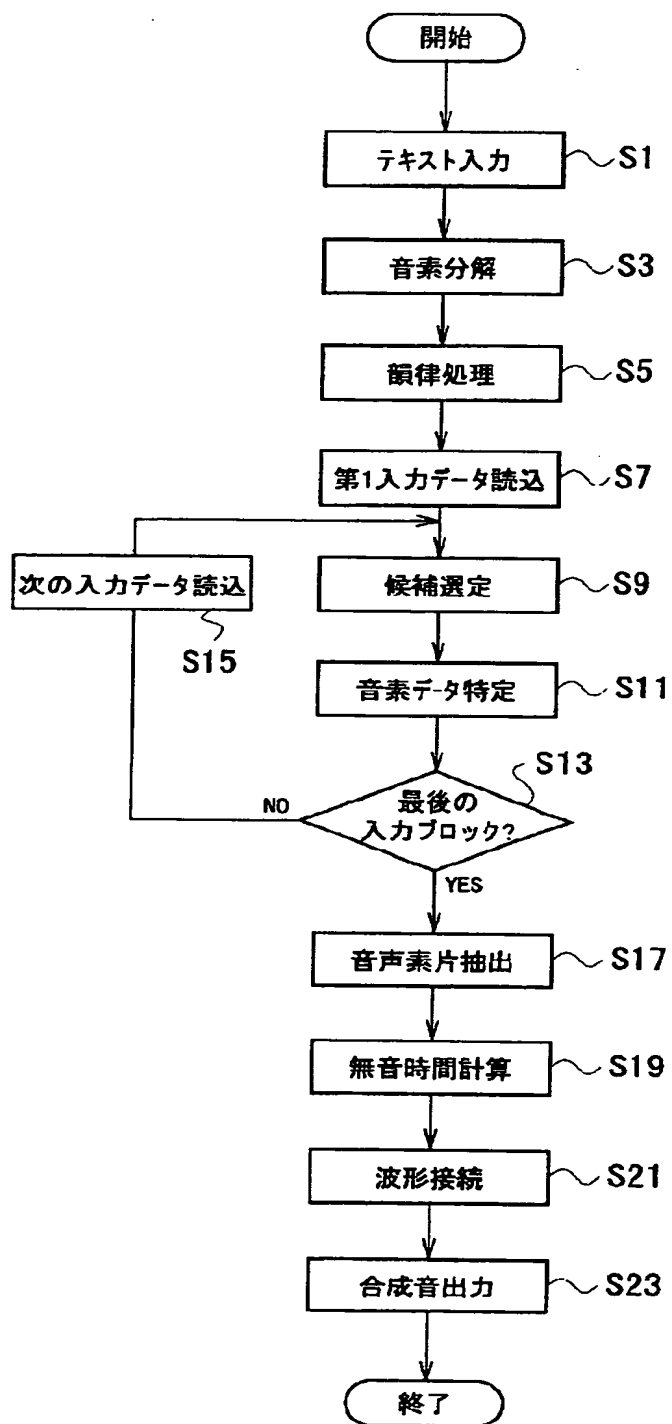
#### 【符号の説明】

- 1 テキスト入力部
- 3 音素分解処理部
- 5 韻律処理部
- 7 音素データ候補選定部
- 9 音素データ特定部
- 11 音声素片抽出部
- 13 無音時間計算部
- 15 音声合成部
- 17 実音声波形記憶部
- 19 音素データ記憶部
- 21 タイミング情報生成部
- 21a 再生時間演算部

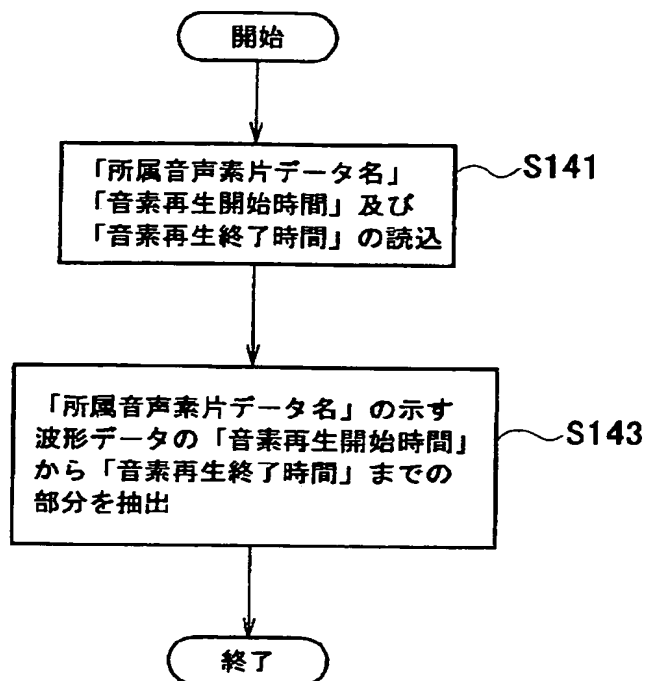
【図 1】



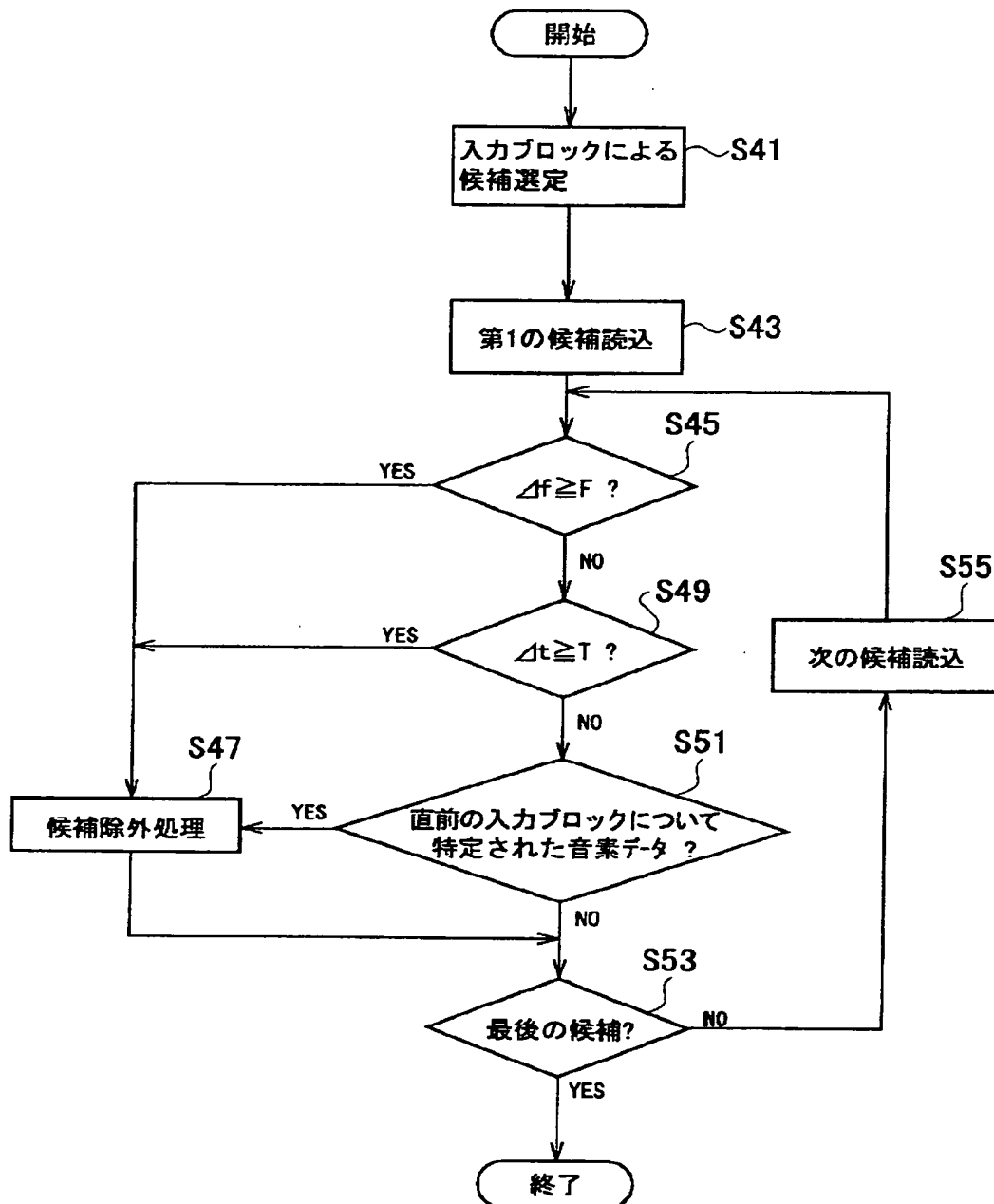
【図2】



【図8】

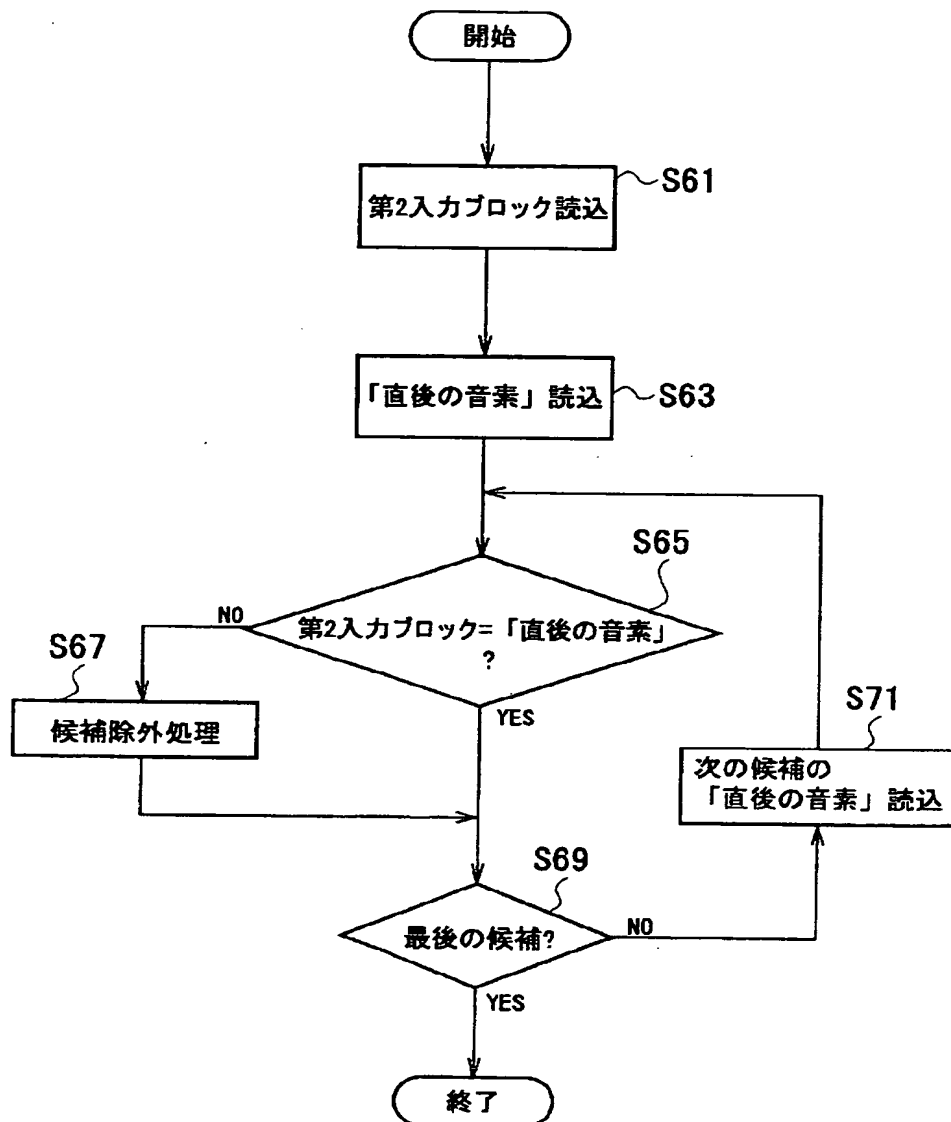


【図3】

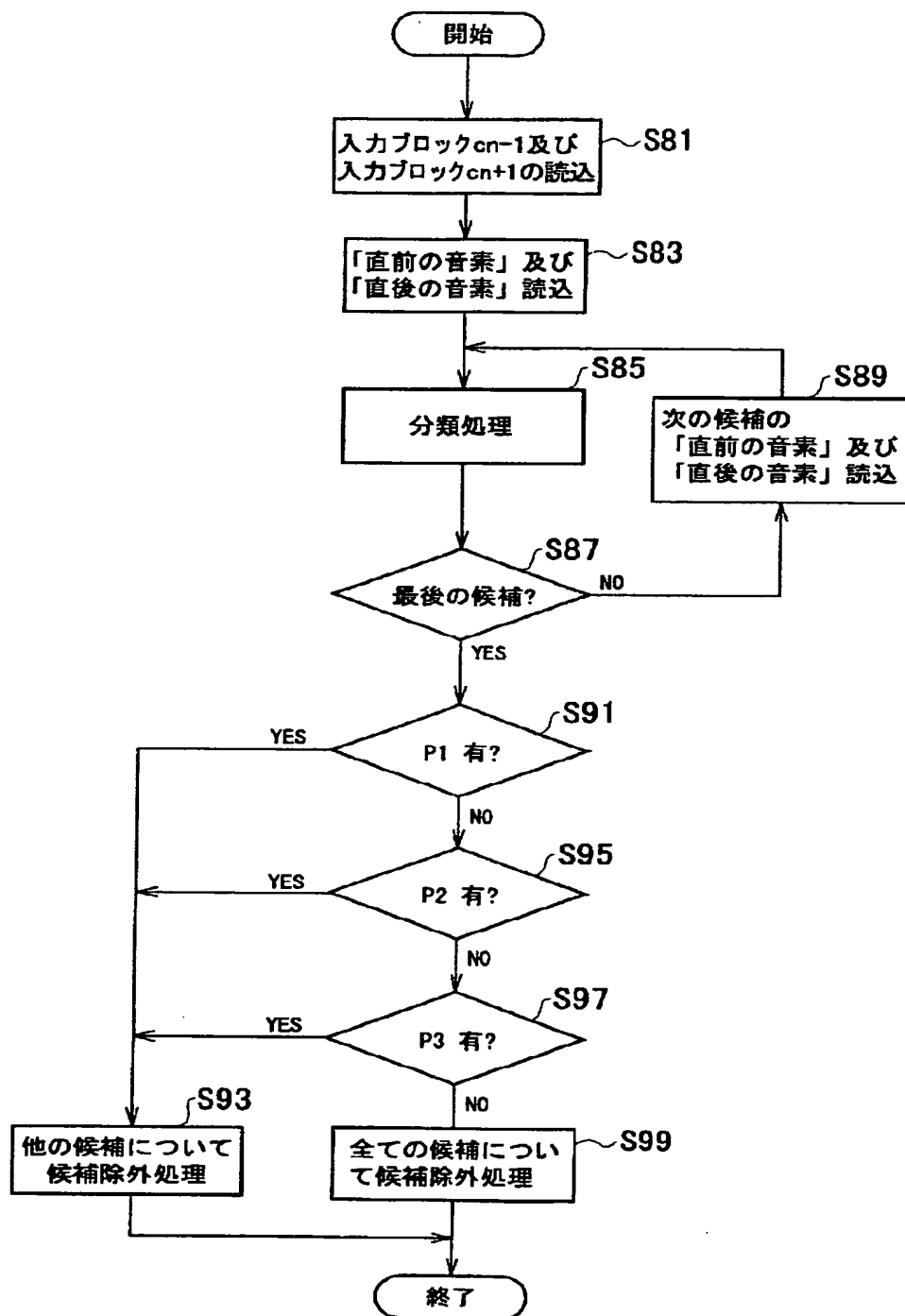




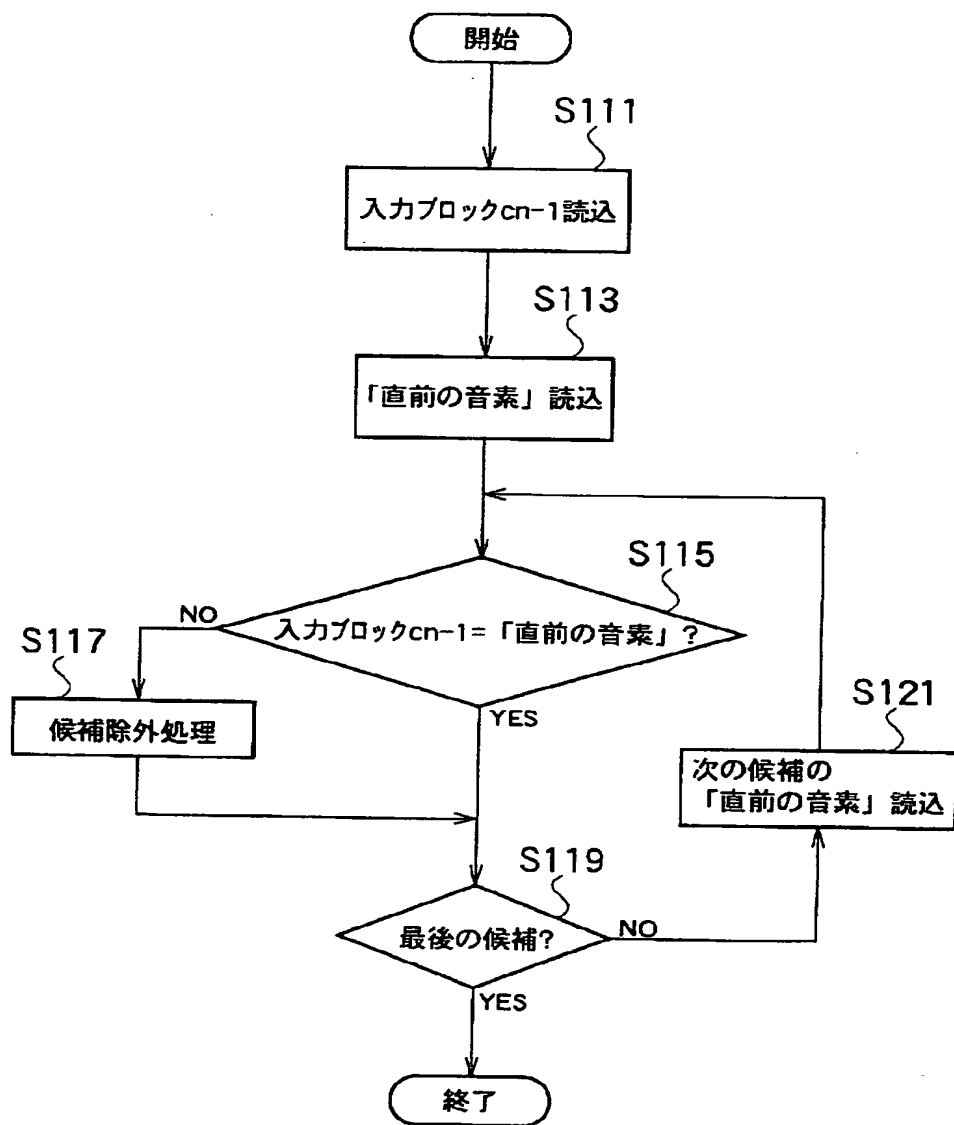
【図4】



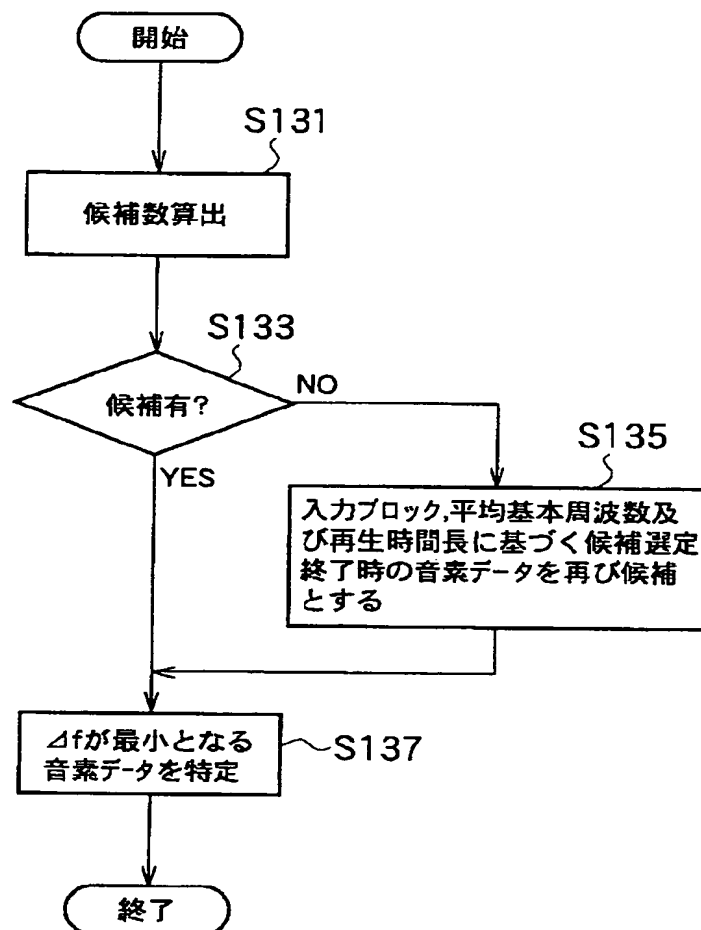
【図5】



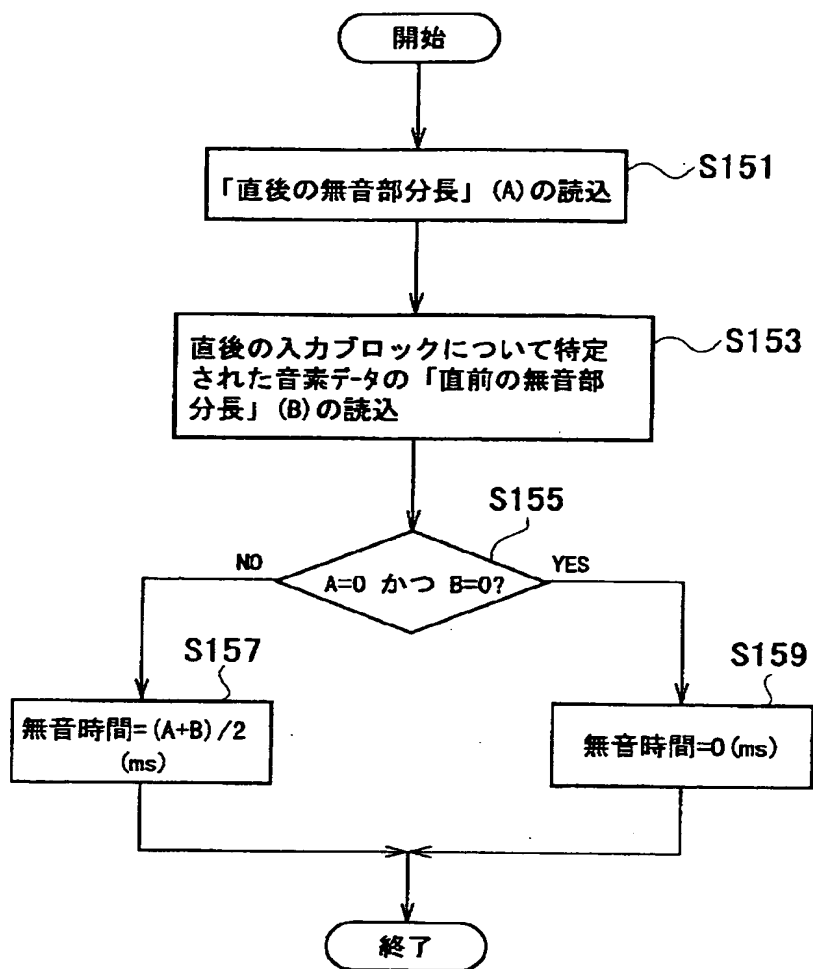
【図6】



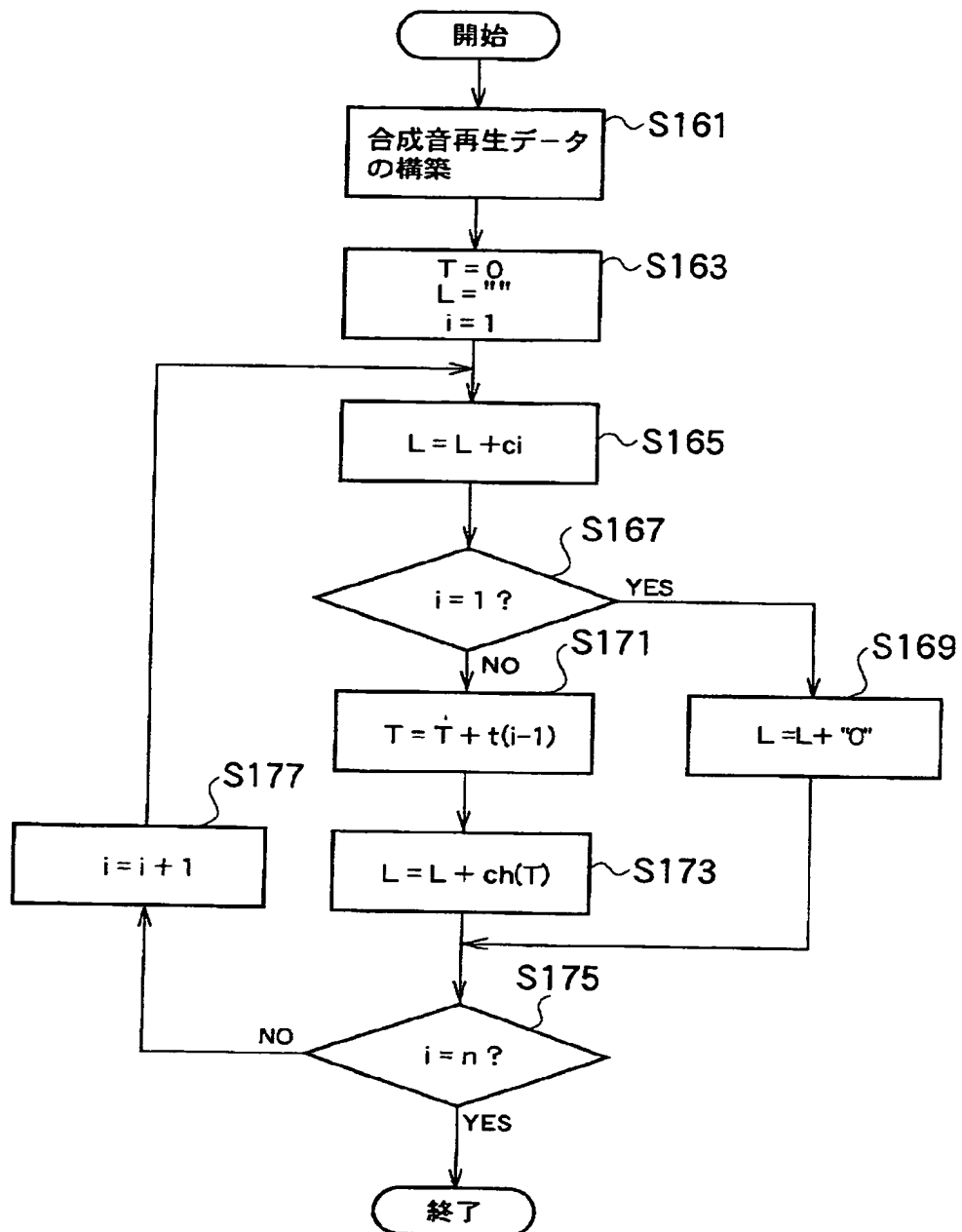
【図7】



【図9】



【図10】



【図11】

